Large-Scale Kinetic Parameters Estimation of Metabolic Model of *Escherichia Coli*

Mohammed Adam Kunna Azrag, Tuty Asmawaty Abdul Kadir, Muhammad Nomani Kabir, and Aqeel S. Jaber

***Abstract*—In the last few decades, the metabolic model of *E.coli* has attracted the attention of many researchers in the area of biological system modeling. Metabolic models are constructed using mass-balance equations with kinetic-rate computation to simulate the behavior of the metabolic system over time. However, in the development of the metabolic model, large-scale kinetic parameters affect the model response if the parameter values are not assigned accurately, which, in turn, propagates the errors in the *ordinary differential equations* (ODEs) – the mass balance equations associated with the model. This situation emphasizes the need to adopt a global optimization technique to compute the kinetic parameters such that the errors – the discrepancy between actual biological data and the model response - are minimized. In this work, the PSO algorithm has been adopted to estimate the kinetic parameters by minimizing the errors of the large-scale of metabolic model response of *E. coli* with reference to real experimental data. Seven highly sensitive kinetic parameters in the model response were considered in the optimization problem. Estimation of the 7th kinetic parameters by the PSO method provides a good performance of the model in terms of accuracy.**

***Index Terms*—Kinetic parameters, dynamic metabolic model, *escherichia coli*, PSO algorithm.**

I. Introduction

In the recent years, the development of metabolic models is the core of system biology. These metabolic models were built using (metabolites, enzymes, and kinetic parameters values) to simulate the behavior of the system [1]. Moreover, in system biology, kinetics computational modeling plays important roles in the analysis of the metabolic process using ordinary differential equations (ODEs) [2]. However, kinetic models depend on the metabolites, enzymes, co-factors, and the number of kinetic parameters values; but due to the large-scale kinetic parameters, this may misbehave if the values are not accurate and the system is nonlinear. This situation requires a second step in investigating the model response after the initial build to determine how much the model is simulating the behavior of the system which caused the model response error changes [3].

More recently, many kinetic models have been presented to simulate the *E. coli* pathways system such as to simulate the pathways of glycolysis and pentose phosphate [2], [4]. Ohno formulated the TCA cycle in Dictyosteliumdiscoideum [5], while [6] integrated the large-scale kinetic parameters with the TCA cycle, acetate formation, and anaplerotic pathway. Lee [7] integrated them with amino acid biosynthesis. Moreover, the kinetic models contain large pathways that need large kinetic parameters, which have been used to detect the concentration changes in the model response. These kinetic models, mostly, are nonlinear and the task of kinetic parameters estimation is a hard problem due to identifiability, interdependence among parameters, and poor data quality [8]; where the metabolic model's response errors in system biology are called parameter estimation problem. This problem is mostly solved by sensitivity analysis and global optimization algorithm [3].

Moreover, recent articles have studied the estimation of large-scale kinetic parameters [4] and proposed the method of stepwise internalization to analyze the sensitivity of kinetic parameters in these glycolysis and pentose phosphate pathways and then applied Simulating Annealing to estimate the kinetics note that contains 85 kinetic parameters. Other authors used Monte Carlo simulation and Sobol method to address the sensitivity. These studies stated that nine kinetic parameters were highly affecting the model response of Embden-Meyerhof, pentose phosphate, and phosphotransferase system. Yet, they formulated dynamic parameter estimation problem to estimate the kinetic parameters sensitivity result using Control Vector Parameterization Approach [8]. The dynamic recursive estimator has been used for the estimation of six parameters and was applied to the model heat shock response in *E. coli* and the model of a synthesis gene regulation system. Lillacci & Khammash stated that the method of state extension allows simultaneous estimation of both metabolites/enzymes and kinetics during the process under investigation [9]. Tosato increased each kinetic parameter up to 200% to investigate the sensitivity analysis of 100 kinetic parameters where seven kinetic parameters of v^{max} are stated as the most effective kinetics. Then they applied Real-Coded Genetic algorithm to that kinetics to be estimated [22]. Similarly, Kunna stated that 7 kinetic parameters affected highly in the model formulated by Kadir *et al*., using the methods of one-at-a-time sensitivity analysis and applied PSO algorithm with four data-set were modified and adopted to give a precise result [10]. Chong *et al*., studied the production of desired metabolites such like G3P (Glyceraldehyde 3-phosphate) and AcO (Acetyl coenzyme A) by proposing an Improved Bee Memory Differential

Manuscript received June 21, 2018; revised March 19, 2019.

Mohammed Adam Kunna Azrag, Tuty Asmawaty Abdul Kadir, Muhammad Nomani Kabir, and Aqeel S. Jaber are with Universiti Malaysia Pahang, Malaysia (e-mail: mohammed87kunna@gmail.com).

doi: 10.18178/ijmlc.2019.9.2.781

Evolution algorithm (IBMDE) through the kinetic parameters estimation of each metabolites and compared to Simulating Annealing, Genetic Algorithm, Differential Evolution and Nelder Mead algorithms [11]. The proposed algorithm of (IBMDE) proved that is sufficient than the others algorithms but not compared to PSO algorithm [11]. Gonzalez *et al.*, they proposed a constructed perturbation function to estimate the parameters of cadBA S-systems in E. coli using Simulating Annealing algorithm [12] were 5 data set are used.

Other researchers investigated the performance of GA and PSO algorithms in parameters estimation of microbial growth using 8 ODEs and 31 parameters executed by ODE15s numerical solver in Matlab platform where the performance of PSO better than GA in estimation [13]. Similarly, [14] stated that four measurements of computational time, aver age of errors rate, stander deviation and, production graph are used to compare the performance of SA, PSO, and downhill simplex methods in parameters estimation applied to the model of essential amino acid production. They proved that, PSO has constancy toward the parameters estimation with reasonable time.

In any case, the PSO algorithm was formulated based on the inspiration of natural behavior of animal foraging activity. These particles such as bird flocking and fish schooling do not have any leader in their group or swarm, [15]. They mention that the PSO algorithm is an efficient method when applied to nonlinear estimation problem [16]. However, [17], [18] they applied PSO, GA, and DE algorithms to estimate the PID controller parameters stated the PSO has constant convergence toward the optimal solution in both linear/non-linear estimations. One of the advantages of PSO algorithm is her ability to improve the global and local exploration abilities. Moreover, the PSO algorithm has proved to be effective in minimizing the steady-state errors [19], [20].

In this study, the model formulated in 2010 by [6] was used as a benchmark to estimate large scale-scale kinetic parameter and minimize the model responses based on the sensitivity analysis result of [10] and the adoption of PSO is proposed to solve the large-scale kinetic parameters estimation of *E. coli* model. This model contains 172 kinetic parameters distributed in five pathways which are glycolysis, pentose phosphate, TCA cycle, gluconeogenesis, glyoxylate pathways in addition to the phosphotransferase system and acetate formation. Thus, the PSO adoption estimates the kinetic parameters and minimize the model under study response errors. Furthermore, the kinetic parameters sensitivity analysis result of [10] was estimated by adopting a PSO algorithm using real experimental data taken from [4].

However, this adoption effectively minimized the model response errors and estimated the kinetic parameters within a reasonable time. The rest of this research paper is structured in the following way: the second section gives a brief description of the dynamic metabolic model of *E. coli* structure; the third section discusses the kinetic parameters that need to be estimated and the PSO adoption algorithm for large-scale estimation; section four presents the result, section five presents the result analyzation, and section six presents the study conclusion.

II. Model Description

The main dynamic metabolic model of *E. coli* formulated [6] was considered as a case study. This model has 23 metabolites, 28 enzymatic reactions with 10 co-factors (e.g., atp, coa, nadhp), continuous culture at steady state with dilution rate 0.1 distributed in five pathways of (glycolysis, pentose phosphate, TCA cycle, gluconeogenesis, glyoxylate in addition to phosphotransferase system and acetate formation) and described in Fig. 1.

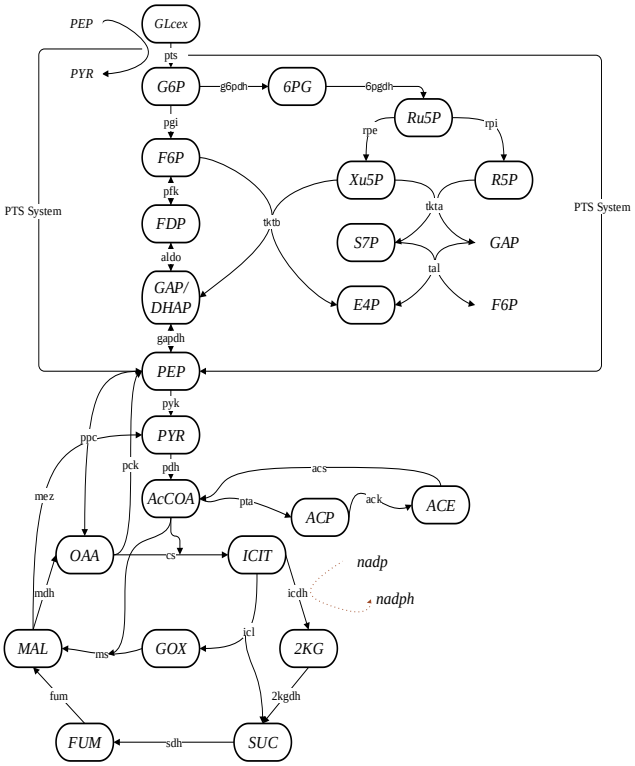

Fig. 1. The main dynamic Metabolic of *E. coli.*

The concentration metabolites rate of the changes in the above model is given by the following equation:

$$\frac{dc_i}{dt} = \sum_j R_{ij} v_i - \mu C_i \tag{1}$$

where c_i is the concentration of metabolite i, R_{ij} is the stoichiometric coefficient of metabolite i in the reaction j, v_i is the rate of the reaction j and μC_i is the growth rate on the dilution effect. However, all the formulas of the kinetic rate equations and the mass balance in this dynamic model are taken from [6] which described in Table I and Table II respectively. Parameter estimation is an essential process of fitting the model into the experimental data, and this requires analysis on the model parameters with the error criterion, defined as the sum of absolutes of difference between experimental and simulated data. Due to the large-scale kinetic parameters involved which affected on the model response under study, the Particle Swarm Optimization algorithm is proposed to identify the kinetics upon stated by the sensitivity analysis kinetics result.

TABLE I: KINETIC RATE EQUATIONS

Reaction's	Kinetic equations
Cell growth	$\begin{cases} \mu_m\left(1-\frac{[X]}{X_m}\right)\left(\frac{[GLc^{ex}]}{K_s+[GLc^{ex}]}\right)k_{ATP}v_{ATP}(.), ([GLc^{ex}]>0) \\ \frac{\mu_{mA}[Ace^{ex}]}{K_{sA}+[Ace^{ex}]}\,k_{ATP}v_{ATP}(.), ([GLc^{ex}]\le 1\ and [Ace^{ex}]>0) \end{cases}$
PTS	$\frac{v_{PTS}^{max}[GLc^{ex}]\frac{[PEP]}{PYR}}{\left(K_{a1}+K_{a2}\frac{[PEP]}{[PYR]}+K_{a3}[GLc^{ex}]+[GLc^{ex}]\frac{[PEP]}{[PYR]}\right)\left(1+\frac{[G6P]^n G6P}{K_{G6P}}\right)}$
PGI	$\frac{v_{PGI}^{max}\left([G6P]-\frac{[F6P]}{K_{eq}}\right)}{K_{G6P}\left(1+\frac{[F6P]}{K_{F6P}\left(1+\frac{[F6P]}{K_{6pgin\ h}^{F6P}}\right)}+\frac{[6PG]}{K_{6pgin\ h}^{G6P}}\right)+G6P}$
PFK	$\frac{v_{PFK}^{max}K_{ATP}[F6P]}{K_{(ATP,ADP)}\left([F6P]+K_z^{F6P}\frac{K_{b(ADP,AMP)}+\frac{[PEP]}{K_{PEP}}}{K_{a(ADP,AMP)}}\right)\left(1+\frac{L_{pfk}}{\left(1+[F6P]\left(\frac{K_{a(ADP,AMP)}}{K_s^{F6P}\left(K_{b(ADP,AMP)}+\frac{[PEP]}{K_{PEP}}\right)}\right)\right)^{n_{PFK}}}\right)}$
Aldo	$\frac{v_{ALDO}^{max}\left([FDP]-\frac{[DHAP][GAP]}{K_{eq}}\right)}{\left(K_{FDP}+[FDP]+\frac{K_{GAP}[DAHP]}{[K_{eq}V_{bif}]}+\frac{K_{DHAP}[GAP]}{[K_{eq}V_{bif}]}+\frac{[FDP][GAP]}{K_{in\,h}^{PEP}}+\frac{[DHAP][GAP]}{K_{eq}V_{bif}}\right)}$
GAPDH	$\frac{v_{GAPDH}^{max}\left([GAP]-\frac{[PEP][NADH]}{K_{eq}[NAD]}\right)}{\left(K_{GAP}\left(1+\frac{[PEP]}{K_{PGP}}\right)+[GAP]\right)\left(\frac{K_{NAD}}{NAD}\left(1+\frac{[NADH]}{K_{NADH}}\right)+1\right)}$
PYK	$\frac{v_{PYK}^{max}[PEP]\left(\frac{PEP}{K_{PEP}}+1\right)^{n_{pyk}-1}[ADP]}{K_{PEP}\left(L_{PYK}\left(\frac{1+\frac{[ATP]}{K_{ATP}}}{\frac{[FDP]}{K_{FDP}}+\frac{[AMP]}{K_{AMP}}+1}\right)^{n_{pyk}}+\left(\frac{[PEP]}{K_{PEP}}+1\right)^{n_{pyk}}\right)([ADP]+K_{ADP})}$
Ppc	$\frac{K_1+K_2[AcCOA]+K_3[FDP]+K_4[AcCOA][FDP]}{1+K_5[AcCOA]+K_6[FDP]}\left(\frac{[PEP]}{K_m+[PEP]}\right)$
G6PDH	$\frac{v_{G6PDH}^{max}[G6P][NADP]}{([G6P]+K_{g6p})\left(1+\frac{[NADPH]}{K_{ndp\ h}}\right)\left(K_{nadp}\left(1+\frac{[NADPH]}{K_{nadp\ h}}\right)+NADP\right)}$
PGDH	$\frac{v_{PGDH}^{max}[6PG][NADP]}{([6PG]+K_{6pg})\left([NADP]+K_{nadp}\left(1+\frac{[NADPH]}{K_{nadp\ h}}\right)\left(1+\frac{[ATP]}{K_{atp}}\right)\right)}$
Rpe	$v_{Rpe}^{max}\left([Ru5P]-\frac{[R5P]}{K_{eq}^{Rpe}}\right)$
Rpi	$v_{Rpi}^{max}\left([Ru5P]-\frac{[R5P]}{K_{eq}^{Rpi}}\right)$
TktA	$v_{TKtA}^{max}\left([R5P][Xu5P]-\frac{[S7P][GAP]}{K_{eq}^{TKtA}}\right)$
TktB	$v_{TKtB}^{max}\left([Xu5P][E4P]-\frac{[F6P][GAP]}{K_{eq}^{TKtB}}\right)$
Tal	$v_{TaL}^{max}\left([GAP][S7P]-\frac{[E4P][F6P]}{K_{eq}^{TKtB}}\right)$
PcK	$v_{PcK}^{max}\left(\frac{[OAA]\frac{[ATP]}{[ADP]}}{K_m^{OAA}\frac{[ATP]}{[ADP]}+[OAA]\frac{[ATP]}{[ADP]}+\frac{K_i^{ATP}K_m^{OAA}}{K_i^{ADP}}+\frac{K_i^{ATP}K_m^{OAA}}{K_m^{PEP}K_i^{ADP}}[PEP]+\frac{K_i^{ATP}K_m^{OAA}}{K_i^{PEP}K_l^{ATP}}\frac{[ATP][PEP]}{[ADP]}+\frac{K_i^{ATP}K_m^{OAA}}{K_i^{ADP}K_l^{OAA}}[OAA]}\right)$
PDH	$\frac{\frac{v_{PDH}^{max}}{[NAD]}\left(\frac{1}{1+K_i\frac{[NADH]}{[NAD]}}\right)\left(\frac{[PYR]}{K_m^{PYR}}\right)\left(\frac{1}{K_m^{NAD}}\right)\left(\frac{[COA]}{K_m^{COA}}\right)}{\left(1+\frac{[PYR]}{K_m^{PYR}}\right)\left(\frac{1}{NAD}+\frac{1}{K_m^{NAD}}+\frac{[NADH]}{K_m^{NADH}[NAD]}\right)\left(1+\frac{[COA]}{K_m^{COA}}+\frac{[AcCOA]}{K_m^{AcCOA}}\right)}$
Pta	$\frac{v_{Pta}^{max}\left(\frac{1}{K_i^{AcCOA}K_m^P}\right)\left([AcCoA][P]-\frac{[AcP][CoA]}{K_{eq}}\right)}{\left(1+\frac{[AcCoA]}{K_i^{AcCoA}}+\frac{[P]}{K_i^P}+\frac{[ACP]}{K_i^{ACP}}+\frac{[CoA]}{K_i^{CoA}}+\left(\frac{[AcCoA][P]}{K_i^{AcCoA}K_m^P}\right)+\left(\frac{[AcP][CoA]}{K_m^{ACP}K_i^{CoA}}\right)\right)}$
Ack	$\frac{v_{Ack}^{max}\left(\frac{1}{K_m^{ADP}K_m^{ACP}}\right)\left([AcP][ADP]-\frac{[ACE][ATP]}{K_{eq}}\right)}{\left(1+\frac{[Acp]}{K_m^{Acp}}+\frac{[ACE]}{K_m^{ACE}}\right)\left(1+\frac{[ADP]}{K_m^{ADP}}+\frac{[ATP]}{K_m^{ATP}}\right)}$
Acs	$\frac{v_{Acs}^{max}[ACE][NADP]}{(K_m+[ACE])(K_{eq}+[NADP])}$
Cs	$\frac{v_{CS}^{max}[AcCoA][OAA]}{(K_d^{AcCoA}K_m^{OAA}+K_m^{AcCoA}[OAA])+\left([AcCoA]K_m^{OAA}\left(1+\frac{[NADH]}{K_{i1}^{NADH}}\right)\right)+\left([AcCoA][OAA]\left(1+\frac{[NADH]}{K_{i2}^{NDAH}}\right)\right)}$
ICDH	$\frac{[ICDH]\frac{K_f}{K_m^{ICIT}}K_d^{NADP}\left([ICIT]-\frac{[NADH][2KG]}{K_{eq}^{ICDH}[NADP]}\right)}{\left(\frac{1}{[NADP]}+\frac{[ICIT]K_m^{NADP}}{K_m^{ICIT}K_d^{NADP}[NADP]}+\frac{1}{K_d^{NADP}}+\frac{[ICIT]}{K_m^{ICIT}K_d^{NADP}}+\frac{[ICIT]}{K_d^{ICIT}[NADP]}\frac{[NADPH]K_m^{NADP}}{K_m^{ICIT}K_m^{NADP}}\frac{}{K_{ein\,h}^{NADPH}}+\right.$ $\left.\frac{[NADPH]K_{ekn\,h}^{2KG}}{K_m^{2KG}K_{en\,he}^{NADPH}[NADP]}+\frac{[2KG]K_m^{NADPH}}{K_m^{2KG}K_{en\,he}^{NADPH}[NADP]}+\frac{[2KG]}{K_m^{2KG}}\frac{[NADPH]}{K_{en\,he}^{NADPH}[NADP]}+\frac{[2KG]K_m^{NADPH}}{K_m^{2KG}K_m^{NADPH}}\frac{[NADPH]}{K_{ekn}^{NADP}[NADP]}\right)}$
IcL	$\frac{v_{lcl-f}^{max}\frac{[ICIT]}{K_m^{ICIT}}}{\left(1+\frac{[ICIT]}{K_m^{ICIT}}+\frac{[SUC]}{K_m^{SUC}}+\frac{[PEP]}{K_m^{PEP}}+\frac{[2KG]}{K_m^{2KG}}+\frac{1}{K_l}\right)}$

MS	$\frac{v_{MS}^{max}\ \frac{[GOX]}{K_m^{GOX}}\ \frac{[AcCoA]}{K_m^{AcCoA}} - v_{MS}^{max}\ \frac{[MAL]}{K_m^{MAL}}}{\left(1+\frac{[GOX]}{K_m^{GOX}}+\frac{[MAL]}{K_m^{MAL}}+\left(1+\frac{[AcCoA]}{K_m^{AcCoA}}\right)\right)}$
$KGDH$	$\frac{v_{2KGDH}^{max}\ [aKG][CoA]}{\left(\begin{array}{c}\frac{K_m^{NAD}\ [aKG][CoA]}{[NAD]}+K_m^{CoA}\ [aKG]+K_m^{2KG}[CoA]+[aKG][CoA]+\frac{K_m^{2KG}K_z[aKG][SUC][NADH]}{K_1^{2KG}K_1^{SUC}[NAD]} \\ \frac{K_m^{2KG}K_z[SUC][NADH]}{K_1^{SUC}[NAD]}+\frac{K_m^{NAD}\ [aKG][CoA][NADH]}{K_1^{NADH}[NAD]}+\frac{K_m^{CoA}\ [aKG][SUC]}{K_1^{SUC}}\end{array}\right)}$
SDH	$\frac{v_{SDH\ 1}v_{SDH\ 2}\left([SUC]-\frac{[FUM]}{K_{eq}}\right)}{K_m^{SUC}\ v_{SDH\ 2}+v_{SDH\ 2}[SUC]+\frac{v_{SDH\ 1}[FUM]}{K_{eq}}}$
Fum	$\frac{v_{Fum\ 1}v_{Fum\ 2}\left([FUM]-\frac{[MAL]}{k_{Fum\ eq}}\right)}{K_m^{Fum}\ v_{Fum\ 1}+v_{Fum\ 2}[FUM]+\frac{V_{Fum\ 1}[MAL]}{K_{eq}}}$
Mez	$\frac{v_{Mez}^{max}\ [MAL][NADP]}{(K_{MAL}+[MAL])\ (K_{eq}+[NADP])}$
MDH	$\frac{v_{MDH\ 1}v_{MDH\ 2}\left([MAL]-\frac{[OAA]}{K_{eq}}\right)}{\left(\begin{array}{c}K_1^{NAD}\ K_m^{MAL}\ \frac{v_{MDH\ 2}}{[NAD]}+K_m^{MAL}\ v_{MDH\ 2}+\frac{K_m^{NAD}\ v_{MDH\ 2}[MAL]}{[NAD]}+v_{MDH\ 2}[MAL]+\frac{K_m^{OAA}\ v_{MDH\ 1}[NADH]}{K_{eq}[NAD]}+\frac{K_m^{NADH}\ v_{MDH\ 1}[OAA]}{K_{eq}[NAD]}+ \\ \frac{v_{MDH\ 1}[NADH][OAA]}{K_{eq}[NAD]}+\frac{v_{MDH\ 1}K_m^{OAA}[NADH]}{K_{eq}K_1^{NAD}}+\frac{v_{MDH\ 2}K_m^{NAD}[MAL][OAA]}{K_1^{OAA}[NAD]}+\frac{v_{MDH\ 2}[MAL][NADH]}{K_1^{NAD}} \\ +\frac{v_{MDH\ 1}[MAL][NADH][OAA]}{K_{eq}K_1^{MAL}[NAD]}+\frac{v_{MDH\ 2}[MAL][OAA]}{K_{II}^{OAA}}+\frac{v_{MDH\ 1}[NADH][OAA]}{K_{II}^{NAD}K_{eq}}+\frac{K_1^{NAD}\ v_{MDH\ 2}[MAL][NADH][OAA]}{K_{II}^{NAD}K_m^{OAA}K_1^{NADH}}\end{array}\right)}$

The above kinetic rate equations are related to the mass balance equation of the model under study which described in Table II:

TABLE II: MASS BALANCE DESCRIPTION

Metabolites	Mass balance description
Cell	$\frac{d[X]}{dt}=\mu[X]$
Extra Glucose	$\frac{d[GLC^{ex}]}{dt}=-v_{PTS}[X]$
Glucose-6-phosphate	$\frac{d[G6P]}{dt}=v_{PTS}-v_{PGI}-v_{G6PDH}-\mu[G6P]$
Fructose 6-phospahte	$\frac{d[F6P]}{dt}=v_{PGI}-v_{PFK}+v_{TKTB}+v_{TAL}-\mu[F6P]$
Fructose 1,6-Phosphate	$\frac{d[FDP]}{dt}=v_{PFK}-v_{ALDO}-\mu[FDP]$
Glyceraldehyde 3-phosphate	$\frac{d[GAP]}{dt}=2v_{ALDO}-v_{GAPDH}+v_{TKTA}+v_{TKTB}-v_{TAL}-\mu[GAP]$
Phosphoenol-pyruvate	$\frac{d[PEP]}{dt}=v_{GAPDH}+v_{PCK}-v_{PTS}-v_{PYK}-v_{PPC}-\mu[PEP]$
Pyruvate	$\frac{d[PYR]}{dt}=v_{PYK}+v_{PTS}+v_{MEZ}-v_{PDH}-\mu[PYR]$
Acetyl-CoA	$\frac{d[AcCoA]}{dt}=v_{PDH}+v_{ACS}+v_{CS}-v_{PTA}-\mu[AcCoA]$
Isocitrate	$\frac{d[ICIT]}{dt}=v_{CS}-v_{ICDH}-v_{ICL}-\mu[ICIT]$
2-Keto-D-gluconate	$\frac{d[2KG]}{dt}=v_{ICDH}-v_{2KGDH}-\mu[2KG]$
Succinate	$\frac{d[SUC]}{dt}=v_{2KGDH}+v_{ICL}-v_{SDH}-\mu[SUC]$
Fumrate	$\frac{d[FUM]}{dt}=v_{SDH}-v_{FUM}-\mu[FUM]$
Malate	$\frac{d[MAL]}{dt}=v_{FUM}+v_{MS}-v_{MDH}-v_{MEZ}-\mu[MAL]$
Oxaloacetate	$\frac{d[OAA]}{dt}=v_{MDH}+v_{PPC}-v_{CS}-v_{PCK}-\mu[OAA]$
Glyoxylate	$\frac{d[GOX]}{dt}=v_{ICL}-v_{MS}-\mu[GOX]$
Acetyl phosphate	$\frac{d[ACP]}{dt}=v_{PTA}-v_{ACK}-\mu[ACP]$
Acetate	$\frac{d[ACE^{ex}]}{dt}=(v_{ACK}-v_{ACS})[X]$
6-Phosphogluconolactone	$\frac{d[6PG]}{dt}=v_{G6PDH}-v_{6PGDH}-\mu[6PG]$
Ribose 5-phosphate	$\frac{d[Ru5P]}{dt}=v_{6PGDH}-v_{RPE}-v_{RPI}-\mu[Ru5P]$
Ribulose 5-phosphoenolpyruvate	$\frac{d[R5P]}{dt}=v_{RPI}-v_{TKTA}-\mu[R5P]$
Xylulose 5-phsophate	$\frac{d[Xu5P]}{dt}=v_{RPE}-v_{TKTA}-v_{TKTB}-\mu[Xu5P]$
Sedoheptulose 7-phosphate	$\frac{d[S7P]}{dt}=v_{TKTA}-v_{TAL}-\mu[S7P]$
Erythrose 4-phsophate	$\frac{d[E4P]}{dt}=v_{TAL}-v_{TKTB}-\mu[E4P]$

III. Methodology

The method used in large-scale kinetic parameters estimation for the main metabolic model of *E. coli* are considered in two steps. The first step is to identify which kinetics has high affection on the model response either decreasing or increasing using sensitivity analysis method, were the result of [10] are considered to be used as a benchmark in this study. The second step is to adopt PSO algorithm and find the accurate kinetic parameters of the sensitivity analysis result by using an objective function that minimizes the model response errors based on real experimental data taken from [4].

However, the goal of large-scale kinetic parameters estimation of the dynamic metabolic model is to identify the best set of kinetic parameters values which minimize the model response errors using real experimental data. The objective function for the estimation was formulated using the equation below:

$$F = \left|\left(W_{f1} - W_{z1}\right) + \left(W_{f2} - W_{z2}\right) + \cdots + \left(W_{fy} - W_{zy}\right)\right| \quad (2)$$

Where F is the objective function, W_{zy} is the model response metabolites result for zy model and W_{fy} is the model simulation response result for fy model.

Moreover, [21] stated that the PSO algorithm is inspired by the food searching behavior of fish and birds with their activities in D-dimensional search space, the best individual position of particle i and the best position of the entire swarm depend on the velocity update derived mathematically as follows:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1\left(p_i(t) - X_i(t)\right) + c_2 r_2\left(G_i(t) - Xit\right) \quad (3)$$

$$X_i(t+1) = X_i(t) + v_i(t+1) \quad (4)$$

where v_i is the particle velocity, ω is the inertia weight, p_i is the best position already found by particle i until t time, G_i is the best value so far obtained by any particle i in the population till t time, X_i is the current particle i solution, $c_1 c_2$ are acceleration coefficients toward p_i and G_i. During the particle searching for global optima, the particles take a part of population as its topological neighbors, where the best value is called local best position.

However, the PSO algorithm was formulated in 1995 by Eberhart and Kennedy as a new heuristic method [15], [16]. This method proved that constant movement of the particles toward the solution as the algorithm progresses is vital in the quest for optimal solutions, in most cases.

In order to adopt Particle Swarm Optimization (PSO) algorithm, there is the need to first initialize of the kinetic parameters values for the sensitivity result and sort the values according to the (minimum/maximum) of each kinetics parameters. Then, initialize the model equations and the experimental data concerning the fitness function as stated above. Next, call these subroutines inside the PSO algorithm toward the optimum solution.

In Algorithm 1 below, the process of designing the PSO adoption algorithm is illustrated. However, during the PSO adoption execution, the maximum iteration number is set to 500; the problem dimension (parameters) is set to 7 kinetic parameters, the number of swarms was set to 100 where each swarm searching for 5 times based on the d dimension of the problem randomly and then calculate its objective function F later the smallest result of F will be selected, a linear inertia weight is set to 0.9 [18], the exploitation coefficient$c_1 = 1.5$, the exploration coefficient $c_2 = 0.8$. An earlier study [13], adopted in this study stated that the best result is achieved if $c_1 > c_2$, and r_1 r_2 are random number between 0 and 1.

At this juncture, it is necessary to describe the variant of the PSO algorithm adopted and implemented in the study in the algorithm below:

Algorithm 1: PSO solution

Input,

Initialize PSO parameters n, $iter$, d, ω, $c_1 c_2$, and $r_1 r_2$;

Initialize the d dimension of kinetic parameter numbers with them random primary position;

Initialize the model equations;

From z data account the fitness: $F = \left|\left(W_{f1} - W_{z1}\right) + Wf2 - Wz2 + \ldots + (Wfy - Wzy)\right|$;

Output,

while ($F > 0$)

from each i;

$iter\ i = 1, i++$;

Update the velocity of particles $v_i(t+1) = \omega v_i(t) + c_1 r_1\left(p_i(t) - X_i(t)\right) + c_2 r_2\left(G_i(t) - X_i(t)\right)$;

Update the position of particles $X_i(t+1) = X_i(t) + v_i(t+1)$;

If the fitness $F >$ p_i best, then $G_i \approx solution$ print G_i as the best solution;

If $F > 0$ return step 2 till the criteria is met or found highly solution;

Print the estimated parameters;

End.

The solution steps of the above PSO algorithm are adopted in this study to estimate the large-scale kinetic parameters and minimize the model response are described as follows:

Firstly, input the number of particles n, the maximum $iter$ number of the entire particles, the inertia weight ω, set the dimension of the problem d with respect to the sensitivity analysis result, set the learning factor $c_1 c_2 \approx 4$, set $r_1 r_2$ which are random numbea r between 0 and 1 in (step 1,2).

Secondly, initialize randomly the position of each kinetic parameter based on the lower and upper bound should be found by the PSO adopted calculation as described in step 3 and the model equations (step 4). Then calculate the first objective function using the equation in step 5 where W_{fy} is the model simulation result for y metabolites and W_{fz} is the experimental data for z metabolites (step 5).

Thirdly, if then fitness $F > 0$, the PSO algorithm will then update the velocity and position based on the information gathered from the first calculation of the fitness (step 7, 8, 9, 10 and 11). This fitness will calculate the differences between the experimental data and the simulation result during PSO adoption calculation.

Fourthly, if the fitness was found greater than the personal best position, then PSO set the global best position as the best solution (step 12). If the F is not equal to zero or

close to 0, return to step 3 and alter some parameter values till the system discovers better solutions (step 13).

Fifthly, print out the estimated parameters in step 9 and then end the program in (step 14).

IV. Test Result

The estimation of large-scale kinetic parameters from the metabolic model is a difficult task due to the nonlinearity of the system, where [10] discovered that during the application of the local sensitivity analysis to [6] model, they found that there are seven kinetic parameters that are affecting highly in the model response. These kinetic parameters are v_{max}^{pyk}, n_{pk}, $icdh$, k_{icdh}^{f}, $k_{icdhnadp}^{d}$, $k_{icdhnadp}^{m}$, and v_{max}^{icl} involved in these reaction rate $V_{pyk}, V_{icdh}, and\ V_{icl}$. The seven kinetics are used to minimize the model response errors of the model under study with real experimental data taken from [4], where the data set are Glc, $G6P$, $F6P$, FDP, PEP, PYR, $6PG$, $RU5P$, $XU5P$, $S7P$, $R5P$, and E4P.

As a matter of fact, during the PSO adoption execution, the values of kinetic parameters were set based on their affection on the model under study without stating the big difference between the upper and lower boundaries of each kinetics [22] and then it will be selected randomly with respect to the first objective function calculation. Later on, the updating of the velocity, position and objective function calculation toward the solution will follow the procedure of PSO algorithm adoption stated in the estimation implementation part.

However, the method of PSO adopted in this study minimizes the model response errors. Thus, this minimization increases FDP highly only rather than the other metabolites. This is due to the other metabolites engaged during the calculation such as TCA cycle and Acetate formation; also the affection of these metabolites data where [6] stated that these OAA, PEP, and PYR are affecting in growth may also miss-direct the model simulation results. In any case, the estimated kinetic parameters found in this study are stated below with their upper and lower values in Table III:

TABLE III: Kinetic Estimation

Kinetics	Original	lower	Upper	Kinetic Estimation
v_{max}^{pyk}	1.085	0.9	1.34	1.032
n_{pk}	3	2.5	3.25	2.647
$icdh$	24.421	23.9	24.6	24 .306
k_{icdh}^{f}	289800	289799.4	289800.7	289799.65
$k_{icdhnadp}^{d}$	0.006	0.004	0.04	0.0372
$k_{icdhnadp}^{m}$	0.017	0.009	0.05	0.0482
v_{max}^{icl}	3.8315	3.3315	4.1	3.594

TABLE IV: Metabolites Concentration

Metabolites	Chassagnole 2002	Kadir 2010	Simulation
Glc	0.0556	0.12203	0.1155
$G6P$	3.48	0.12989	0.21931
$F6P$	0.6	0.021457	0.022598
FDP	0.272	1.5186	2.5257
PEP	2.67	1.5076	1.9186
PYR	2.67	2.8279	3.1882
$6PG$	0.808	0.017854	0.01876
$Ru5P$	0.111	0.021398	0.022489
$Xu5P$	0.138	0.026516	0.0803
$S7P$	0.276	0.00473	0.03424
$R5P$	0.398	0.076388	0.027912
$E4P$	0.098	0.027837	0.004318

However, the model simulation response was achieved based on the kinetics estimation by adopting PSO algorithm and compared to the original data set presented by [4] and the model under study response [6] where the simulation result is closer to the experimental data than the model under study. Moreover, the simulated result is described in Table IV.

The analyzation of the result will be described in the next section.

V. Analyzation

The analyzation of the errors minimization shows that 8 out of the 12 datasets are moving toward the real experimental data. Those metabolites are GLx, $G6P$, $F6P$, PEP, $6PG$, $RU5P$, $XU5P$, and $S7P$ while these metabolites of FDP, PYR, $R5P$, and $E4P$ are not minimized will maybe due to the model complexity, lack of data and the lump of some metabolites. Moreover, the $G6P$ experimental data is $3.48\ mM$ and if compared to the $G6P$ model, the simulation result of $G6P$ is moving toward the experimental data rather than the $G6P$ result model under investigation. This difference may be due the model determination of $G6P$ concentration in the presence of $5\ mM\ \ MgSO_4$ and $0.48\ mM\ \ NADP^{+}$ after the addition of $0.7\ U\ ml^{-1}$ of $G6PDH$ reported by [6].

Moreover, the FDP and $E4P$ increased highly. This may be due to the nonlinearity of the system or the lumping of GAP and $DHAP$ metabolites in one equation. Moreover, the changes in FDP may be due to the lumping of pykII and pykI as stated in [6].

As seen in Table IV, the errors minimization simulation response is moving toward the experimental data in 8 metabolites. In addition, the model under investigation has five pathways as compared to the data of only two pathways where four metabolites are not minimized well due to the lack of experimental data.

However, the adoption of the PSO algorithm to minimize the model responses under study was proposed as below with affected result in Fig. 2. The proposed model after the minimization caused changes in some others metabolites either highly decreasing/increasing or small decreasing/increasing if compared to the model under study.

In the glycolysis pathway most of the metabolites are increased highly especially in $G6P$, FDP, $GAP/DHAP$, and PYR this may-be due to the decreasing in Glc^{ex}, the lump of $GAP/DHAP$ metabolites, and the enzymatic affection of $pts\ system$, pfk, $aldo$, $tktb$, and $gapdh$ involvement.

The pentose phosphate pathway almost the metabolites are not affected highly only small changes either increasing or decreasing with perfect result.

The TCA cycle and glycoxylate pathways are affected by highly decreasing in $ICIT$, GOX, $2KG$, SUC, FUM, and MAL metabolites; this may be due to highly increasing in PEP and PYR; where PEP has been affected by 5 enzymatic reaction of $pts\ system$, $gapdh$, pck, ppc and pyk while PYR consuming pyk and mez also may-be due to the involvement of $AcCOA$ in the both pathways.

The gluconeogenesis pathway has been affected and caused highly increasing in PYR and PEP while MAL

decreased highly due to them involvement in glycolysis and TCA cycle pathways. The acetate formation small changes happened either increasing or decreasing.

The calculation of PSO adaptation of large-scale kinetic parameters estimation was performed in Matlab platform and done in 26.3114h where PSO adaptation takes the self-time calculation of 3.346s. As seen in Fig. 2, the ODE function used in this estimation is ODE15s where the condition is the wild type.

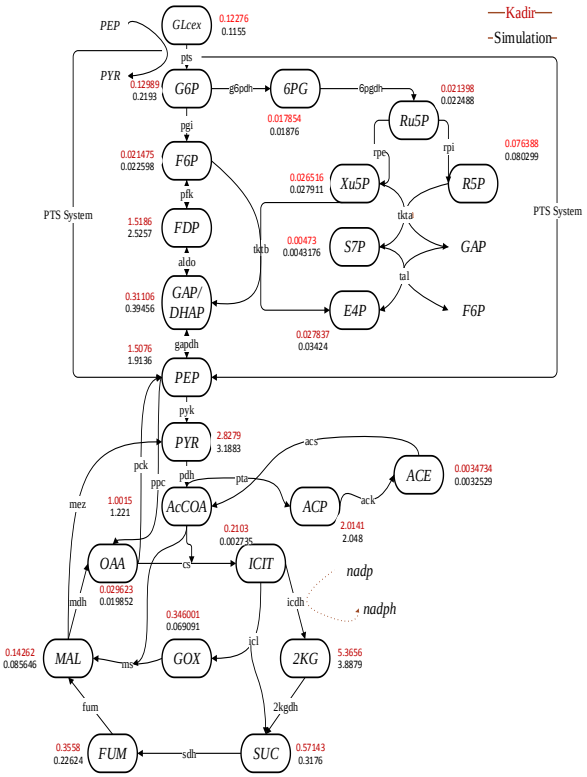

Fig. 2. The model simulation comparison.

After the kinetic parameters are estimated and the model responses under study are minimized, there is increasing and decreasing in the model pathways response simulation result as compared to the model under investigation described in Fig. 3. In the glycolysis pathway the model response simulation of $GLcex$ is decreased while $G6P$, $F6P$, FDP, $GAP/DHAP$, PEP, and PYR are increased maybe due to pts system or small consumption of $GLcex$. In the pentose-phosphate pathway the model response simulation of $6PG$, $Ru5P$, $R5P$, $Xu5P$, and $E4P$ was increased while $S7P$ is decreased maybe due to the increasing in $G6P$ and the involvement of $F6P$ in the calculation. In the TCA cycle and glycoxylate pathways the model response simulation of $ICIT$, OAA, $2KG$, SUC, FUM, MAL, and GOX are decreased maybe due to the gluconeogenesis pathway involvement with affection of mez, pck, and ppc enzymes and the increasing of $PEP$and PYR while decreasing in OAA highly in MAL. Moreover, the Acetate formation has small affection on the model response; which cause increasing in ACP, $AcCOA$ and decreasing in ACE this maybe due to $AcCOA$ affection on the TCA cycle and glyoxylate pathways. However, the increasing was described in green color and the decreasing in black color while the model response under study is marked by red color below in Fig. 3. Finally, the large scale kinetic parameters estimation and the errors minimization of the model under study was achieved with perfect result as stated above.

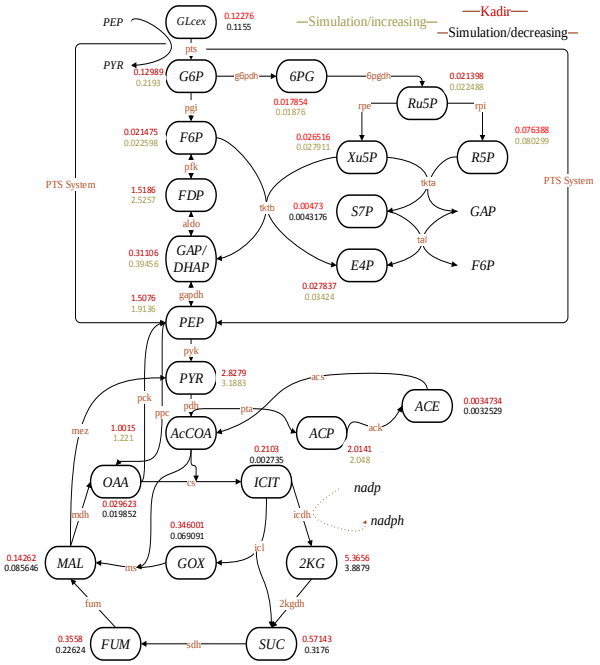

Fig. 3. The increasing/ decreasing of the model response.

VI. Conclusion

The estimation of large-scale kinetic parameters through the minimization discrepancy between the model response and the real experimental data is the primary target in this research. Seven kinetic parameters stated as the most effective parameters in the model response as mentioned in [8] were used in this research. However, the task of estimation is difficult due to the non-linearity of the model. PSO method proved the effectiveness for the estimation of seven parameters with a good level of accuracy within a reasonable time (1 day and 2.3 h). Errors may be more reduced if the further parameters are considered or more experimental data are available. It is hereby recommended that further research investigation is carried out on the estimation of kinetic parameters using more parameters with some newer optimization algorithms like the African Buffalo Optimization, Firefly Algorithm, and Bat Algorithm.

Acknowledgment

Authors appreciate the Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang for funding this research under Grant (RDU 160101).

References

[1] L. P. S. D. Carvalho, S. M. Fischer, J. Marrero, C. Nathan, S. Ehrt, and K. Y. Rhee, "Metabolomics of mycobacterium tuberculosis reveals compartmentalized co-catabolism of carbon substrates," *Chemistry and Biology*, vol. 17, no. 10, pp. 1122-1131, 2010.

[2] M. Breit and M. Netzer, K. M. Weinberger, and C. Baumgartner, "Modeling and classification of kinetic patterns of dynamic metabolic biomarkers in physical activity," *PLoS Computational Biology*, vol. 11, no. 8, 2015.

[3] L. W. Votapka and R. E. Amaro, "Multiscale estimation of binding kinetics using Brownian dynamics, molecular dynamics and milestoning," *PLoS computational Biology*, vol. 11, no. 10, 2015.

[4] C. Chassagnole, J. W. Schmid, K. Mauch *et al*., "Dynamic modeling of the central carbon metabolism of Escherichia coli," *Biotechnology and Bioengineering*, vol. 79, no. 1, pp. 53-73, 2002.
[5] S. Ohno, C. Furusawa, and H. Shimizu, "In silico screening of triple reaction knockout Escherichia coli strains for overproduction of useful metabolites," *Journal of Bioscience and Bioengineering*, vol. 115, no. 2, pp. 221-228, 2013.
[6] T. A. A. Kadir, A. A. Mannan, A. M. Kierzek, J. McFadden, and K. Shimizu, "Modeling and simulation of the main metabolism in Escherichia coli and its several single-gene knockout mutants with experimental verification," *Microbial Cell Factories*, vol. 9, no. 1, 2010.
[7] F. C. Lee, G. P. Rangaiah, and D. Y. Lee, "Modeling and optimization of a multi-product biosynthesis factory for multiple objectives," *Metabolic Engineering*, vol. 12, no. 3, 2010.
[8] J. D. Maggio, C. Paulo, V. Estrada, N. Perotti, J. C. D. Ricci, and M. S. Diaz, "Parameter estimation in kinetic models for large scale biotechnological systems with advanced mathematical programming techniques," *Biochemical Engineering Journal*, vol. 83, pp. 104-115, 2014.
[9] G. Lillacci and M. hammash, "Parameter estimation and model selection in computational biology," *PLoS Computational Biology*, vol. 6, no. 3, 2010.
[10] M. A. Kunna, T. A. A. Kadir, A. S. Jaber, and J. B. Odili, "Large-scale kinetic parameter identification of metabolic network model of e. coli using PSO," *Advances in Bioscience and Biotechnology*, vol. 6, no. 2, 2015.
[11] C. K. Chong, M. S. Mohamad, S. Deris, M. S. Shamsir, L. E. Chai, and Y. W. Choon, "Parameter estimation by using an improved bee memory differential evolution algorithm (IBMDE) to simulate biochemical pathways," *Current Bioinformatics*, vol. 9, no. 1, pp. 65-75, 2014.
[12] O. R. Gonzalez, C. Küper, K. Jung, P. C. Naval, and E. Mendoza, "Parameter estimation using Simulated Annealing for S-system models of biochemical networks," *Bioinformatics*, vol. 23, no. 4, pp. 480-486, 2006.
[13] D. Calçada, A. Rosa, L. C. Duarte, and V. V. Lopes, "Comparison of GA and PSO performance in parameter estimation of microbial growth models: A case-study using experimental data," in *Proc. 2010 IEEE Congress on Evolutionary Computation (CEC)*, 2010, pp. 1-8.
[14] S. T. Ng, C. K. Chong, Y. W. Choon, L. E. Chai, S. Deris, R. Illias, and M. S. Mohamad, "Using particle swarm optimization for estimating kinetics parameters on essential amino acid production of arabidopsis thaliana," *Advances in Biomedical Infrastructure*, pp. 51-61, 2013.
[15] I. D. Boussaï J. Lepagnot, and P. Siarry, "A survey on optimization metaheuristics," *Information Sciences*, vol. 237, pp. 82-117, 2013.
[16] H. Chen, S. Li, J. Liu, F. Liu, and M. Suzuki, "A novel modification of PSO algorithm for SML estimation of DOA," *Sensors*, vol. 16, no. 12, 2016.
[17] W. J. Shen and H. X. Li, "Multi-scale parameter identification of lithium-ion battery electric models using a PSO-LM algorithm," *Energies*, vol. 10, no. 4, 2017.
[18] S. M. Othman, M. F. A. Rahmat, S. M. Rozali, and S. Salleh, "PID parameters optimization using PSO technique for nonlinear electro hydraulic actuator," *Jurnal Teknologi*, vol. 77, no. 28, pp. 67-72, 2015.
[19] J. B. Odili, M. N. M. Kahar, and A. Noraziah, "Parameters-tuning of PID controller for automatic voltage regulators using the African buffalo optimization," *PLoS ONE*, vol. 12, pp. 1-17, 2017.
[20] A. Tarique and H. A. Gabbar," Particle swarm optimization (PSO) based turbine control," *Intelligent Control and Automation*, vol. 4, no. 126, 2013.
[21] C. Li, R. Zhai, and Y. Yang, "Optimization of a heliostat field layout on annual basis using a hybrid algorithm combining particle swarm optimization algorithm and genetic algorithm," *Energies*, vol. 10, no. 11, 2017.
[22] Y. Tohsato, K. Ikuta, A. Shionoya, Y. Mazaki, and M. Ito, "Parameter optimization and sensitivity analysis for large kinetic models using a real-coded genetic algorithm," *Gene*, vol. 518, no. 1, pp. 84-90, 2013.

Mohammed Adam Kunna Azrag obtained his B.Sc. in 2010 in computer studies from the university of National Ribat University and the M.Sc. degree in computer science from the Universiti Malaysia Pahang in 2015. He is currently a PhD candidate in Universiti Malaysia Pahang. His research interests cover the optimization algorithms and data base.

Tuty Asmawaty Abdul Kadir was born in Pahang, Malaysia in 1976. She received the B.Sc. degree (with honors) in computer science and M.Sc. degree in computer science from the University Technology of Malaysia, Malaysia in 1999 and 2005 respectively, and the Ph.D. degree in engineering (Information Science) from the Kyushu Institute of Technology, Japan, in 2010. She currently a senior lecturer at the Universiti Malaysia Pahang, Malaysia, since 2002. Her research interests cover the modelling and simulation, serious games and image processing.

Muhammad Nomani Kabir obtained his MSc in 2003 in computational sciences in engineering and his PhD in 2007 in computer science from the University of Braunschweig, Germany. He is currently a senior lecturer at the Faculty of Computer Systems and Software Engineering, University Malaysia Pahang (UMP), Malaysia. He has been a member of IAENG since 2014. He is working as a project leader and a member on many funded research projects from UMP and national grants. He published over 40 journal and conference papers. His research interests include issues related to information security, modelling and simulation, computational methods, image processing and computer networks.

Aqeel S. Jaber was born in Iraq in 1977. He received the B.E., M.E., degrees from University of Technology, Bagdad, Iraq, in 2001, 2007 respectively, and Ph.D. degree from University Malaysia Pahang, Pahang, Malaysia, in 2015.

He has held lecturing positions at the Department of Electrical Power Engineering Techniques, Al-Mamon University collage, Baghdad, Iraq since 2009. He is currently a senior lecturer at the Al-Mamon University collage.